Centre for Environmental Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

CEDA Data Quick Start User Guide

Dataset Collection:

Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current)

http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0

16/09/2019

Version 1.3

Introduction

This document aims to give additional information to aid use of the data within the full MIDAS land and marine surface station data to accompany information already available within the CEDA data catalogue pages within this collection. This document gives further information including:

- How Station Data are connected in MIDAS (Station source id)
- Extracting data into columns in Excel
- State indicators
- Quality control (QC) guide
- Met element name _j

For information about accessing data, geographic and temporal coverage, please refer to the MIDAS data collection catalogue and the related individual dataset pages using the link above.

How Station Data are connected in MIDAS (Station source id)

The Met Office collates data from a number of observation networks both in the UK and internationally for various purposes, including storing these data in their long term database – MIDAS. As a station may issue meteorological data within a variety of networks, within which it may have a different identifier (e.g. WMO station ID), the Met Office allocate a specific station ID for all data for that station – known as the src_id. The Met Office also allocates an Id type for identifying a recording instrument and a message type giving an indication of the range observations that are recorded.

Source id (src_id) is an integer running from 1 upwards unique to all stations. All observations in the database are stored with their source identifier.

Id_type identifies a recording instrument, e.g. a raingauge (RAIN), an anemometer (WIND), a range of weather recording instruments at a UK Met Office Station (DCNN), a range of recording instruments at an airport (ICAO), a range of recording instruments at a WMO station (WMO), etc.

Met Domain or message types give an indication of the range of observations that can potentially be recorded at the station. For example, WADRAIN is the message type or Met Domain which will be used to send out daily rainfall data as observed at the station of interest from a station operated by a Water Authority (hence the WA prefix). The distinctive message types indicate that the stations are reporting within different networks and each may include different parameter set.

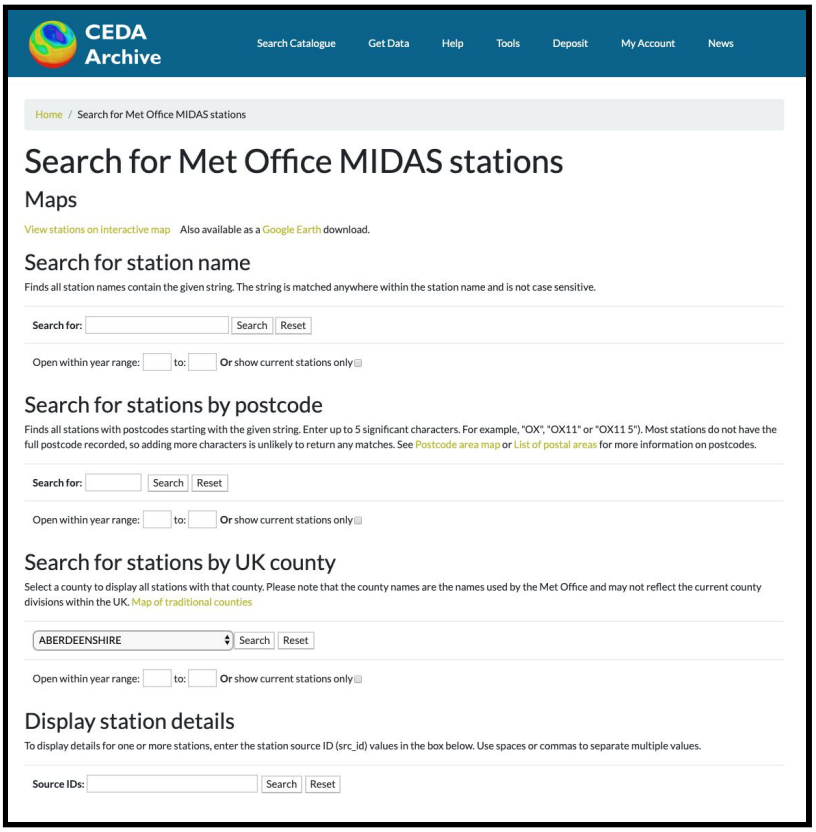
This document shows where all the elements from the SYNOP, NCM, HCM, SREW, METAR and DLY3208 (daily climate) message types are stored in MIDAS. One notable exception is the marine met domains (SHIP, RIG etc.) - all of these met domains have their data stored in the marine_ob (and possibly marine_ice_ob) table only.

Finding relevant stations - MIDAS Station Search Tool

To help find a relevant station or to see detailed station details within MIDAS, CEDA have produced the MIDAS Station Search tool:

http://archive.ceda.ac.uk/midas_stations/

This allows a user to search for a station by name, postcode, county or station 'src_id'. Additionally, the service links to an interactive map and Google Earth files to give additional search options.. The table below shows the strengths and weaknesses of using each method.

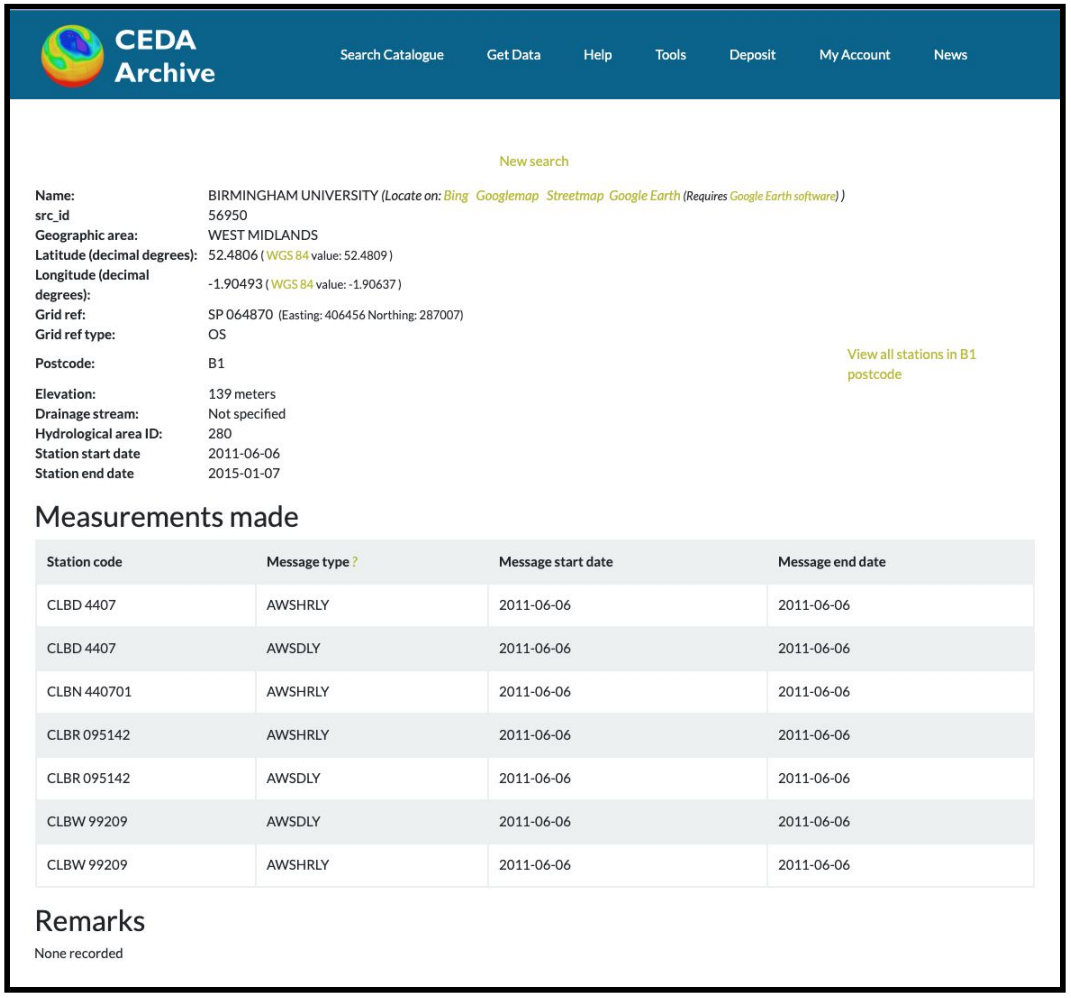


Search option	Strengths	Weaknesses
Name	Likely to match against a name for the area	Some stations may be given a different name within MIDAS than would be expected.
Postcode	All the stations in the postcode area will appear in the options list.	This will miss nearby stations in neighbouring postcode areas
County	All the stations in the county will appear in the options list.	**The Met Office has used historic county borders, therefore places may appear in different counties. The link below shows the county boundaries used.**

		http://archive.ceda.ac.uk/midas_stations/traditional_counties_map.html. Limited to UK coverage only.
Interactive Map	All the stations in the location can be seen on a map. Stations can be searched by both message types and geographical areas. Non-UK stations also displayed.	The stations will not appear in a list, they will appear on a map.
Google Earth	All the stations in a location can be seen on a map. The stations can be searched by message type or geographical area seen on the map in different colours. Non-UK stations also displayed.	Files have to be downloaded according to your search requirements and required Google Earth. The stations will not appear in a list, they will appear on a map.

Please note that not every station reports the same types of data as the instrument in operation will vary from site to site. The MIDAS system will bring together all the data from a particular station regardless of which network it is reporting on.

Once the relevant station has been found, click on it to display more information. The record below shows the information about Birmingham University (src_id: 56950).



A list of the different message types and descriptions can be found in the link below.

http://artefacts.ceda.ac.uk/badc_datadocs/ukmo-midas/met_domain.html

The message types are important as they represent each type of observation and measurements that are taken at the different stations. For example, Birmingham University (shown above) has message type AWSHRLY, which means it is an Automatic Weather Station recording hourly rainfall, temperature and weather observations.

Extracting data for stations

The MIDAS data can be downloaded directly as yearly files, but as these files contain all station data for the given type (e.g. these files are too large for most users therefore a guide has been produced. To extract the data CEDA has developed a tool within the CEDA Web Processing Service (WPS), by using this link below, the MIDAS data can be extracted for one or more weather stations at a time. **This process is strongly advised as there are data from a large number of stations within the MIDAS collection.**

http://wps-web1.ceda.ac.uk/ui/home

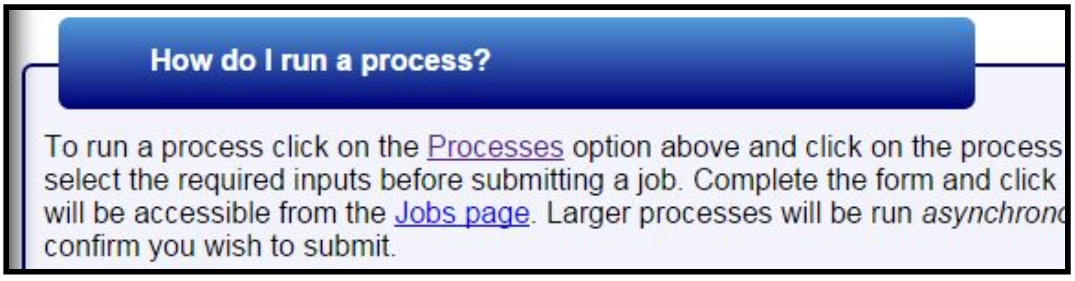
Guidelines on how to use the CEDA WPS to extract MIDAS data is available at:

http://artefacts.ceda.ac.uk/badc_datadocs/ukmo-midas/WPS.html

A step by step guide

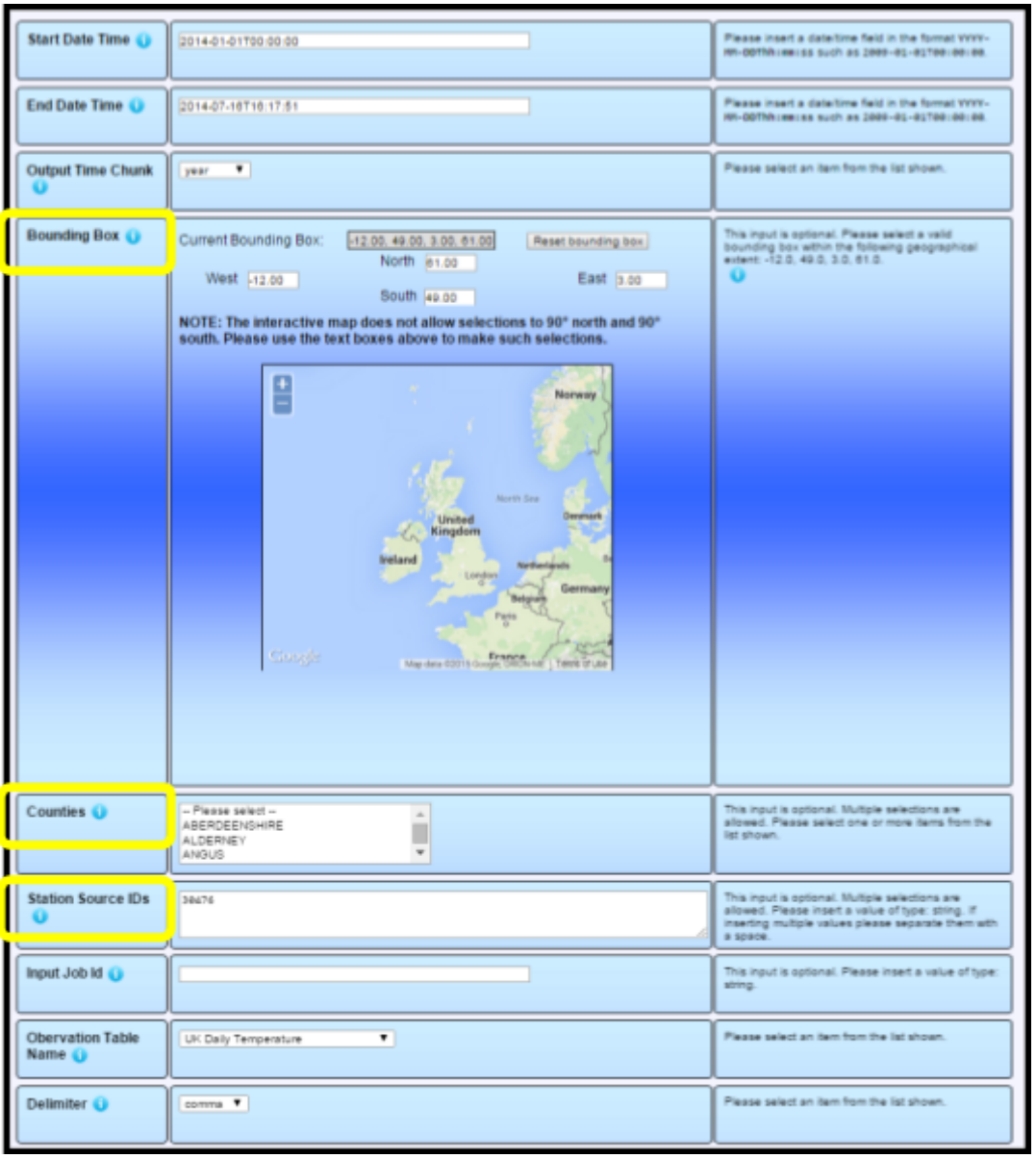
1. Login in

2. Select processes

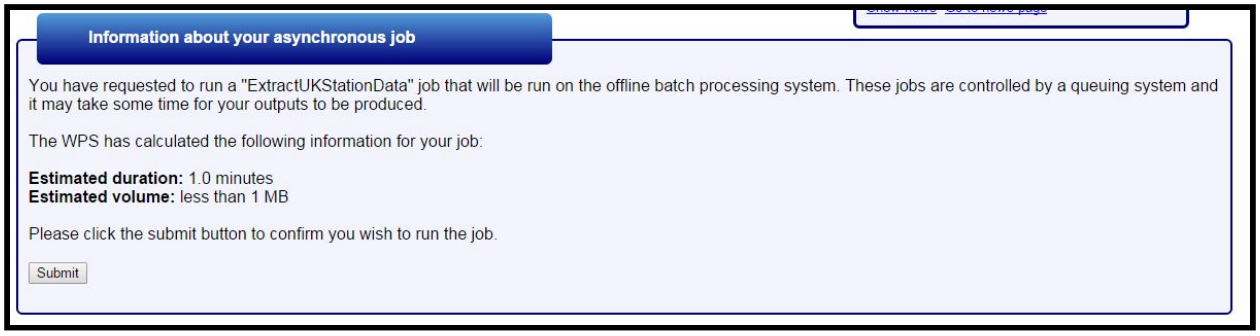


3. Click on extract UK Station Data

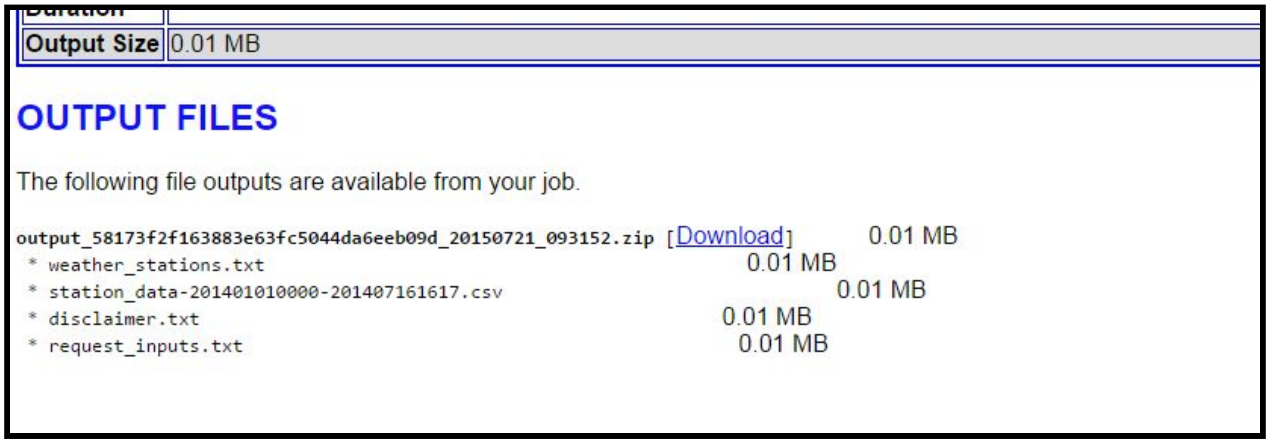
4. Fill in the following fields with the appropriate parameters. It is optional to fill in bounding box, counties and station source ids, but one option must be filled in as highlighted in the diagram below.



5. Press submit and an estimated time and volume is displayed



6. Then the data can be downloaded.



These links below define the column headings depending on which data has been extracted.

Data extracted	Webpage to the relevant column headings
UK Daily Temperature	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo-midas/TD_Table.html
UK Soil Temperature	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo-midas/ST_Table.html
UK Daily Rainfall data	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo-midas/RD_Table.html
UK Daily Weather Observation data	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo-midas/WD_Table.html
UK Hourly Rainfall data	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo-midas/RH_Table.html
UK Hourly Weather Observation data	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo-midas/WH_Table.html

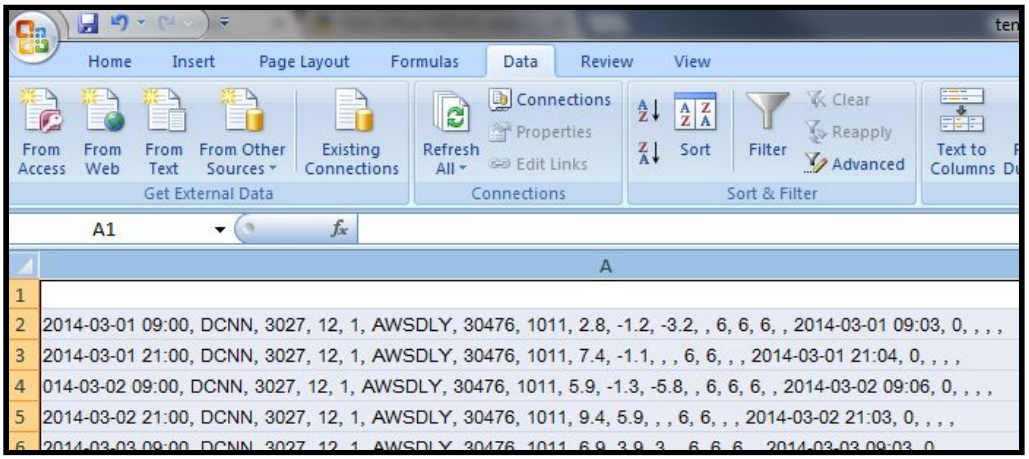
UK Mean Wind data	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo-midas/WM_Table.html
UK Soil Minimum Temperatures (1959-1970 only)	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo-midas/TMSL_Table.html
UK Sub- hourly Rainfall data only to April 2005	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo-midas/RS_Table.html
Global Radiation Observations	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo-midas/RO_Table.html

Handling extracted data

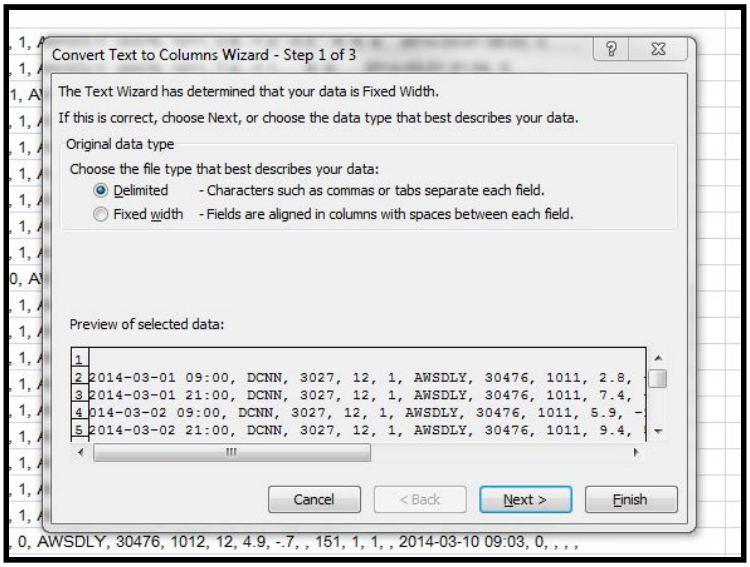
Extracting data into columns in Excel

The link below demonstrates how to convert a text file to Excel

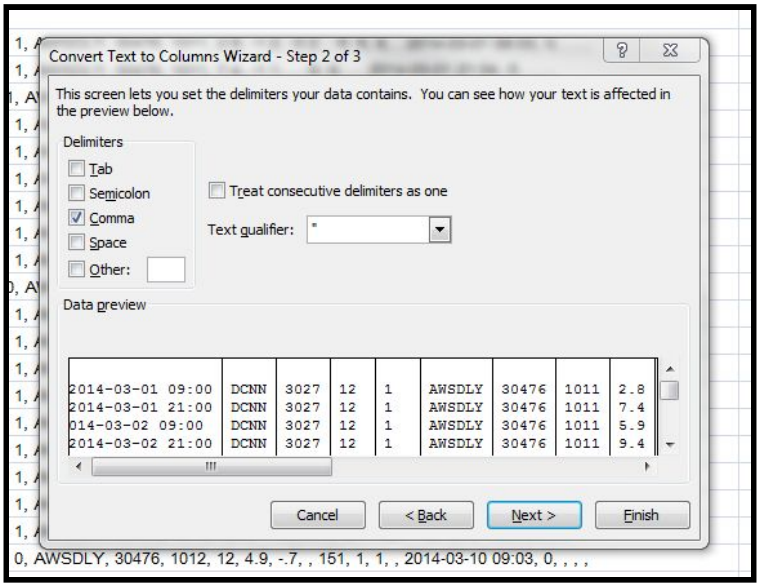
1. Highlight all the data
2. Click on the data tab and into Text to columns



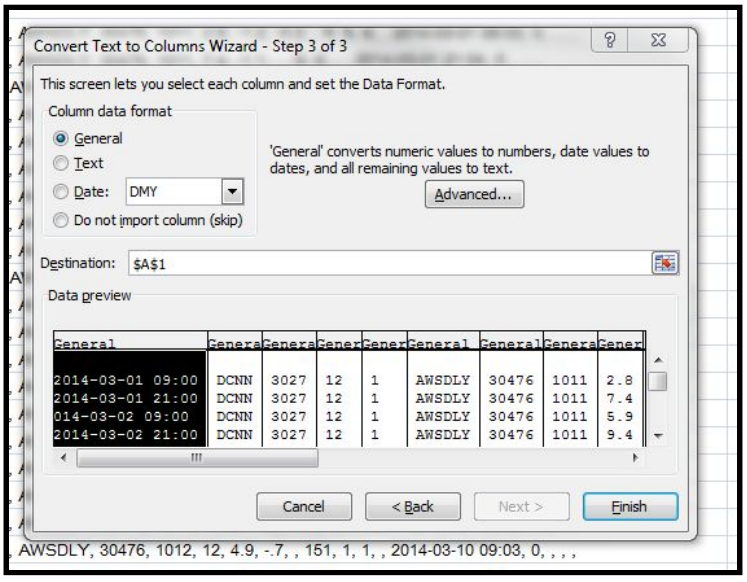
3. Choose the 'Delineated' option



4. Choose a delimiter to be ‘comma’



5. Select finish



Cleaning up the data

At first glance there may be more than 1 entry for a given dataline for a given station’s report type and timestamp. This is due to a couple of factors which need to be accounted for:

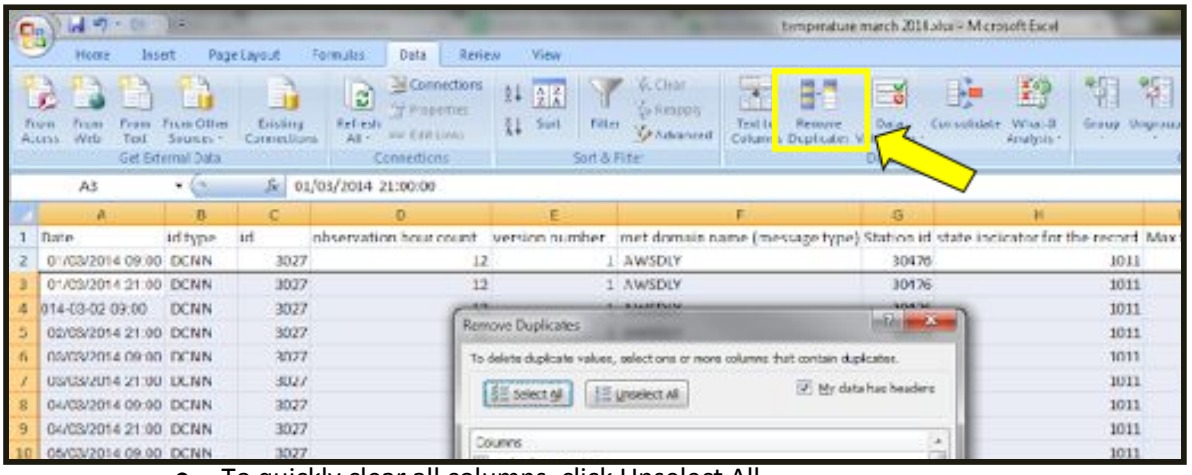
1. When data are initially received by the Met Office’s system they are noted as ‘version 1’ of the data
2. Subsequently, if the entry is to be updated, e.g. following quality control, the *original* line is retained with a version number of zero and the new entry becomes the new version 1 line.
3. Within the Met Office’s system subsequent changes to the data line will then replace the version 1 entry only, meaning that intermediate changes are not stored by the MIDAS system.
4. However, as CEDA’s copy is obtained by periodic snapshots extracting the previous 12 months’ worth of data and only entirely duplicate lines are removed, it is possible that the

intermediate version 1 record states are captured where changes are more than a month apart.

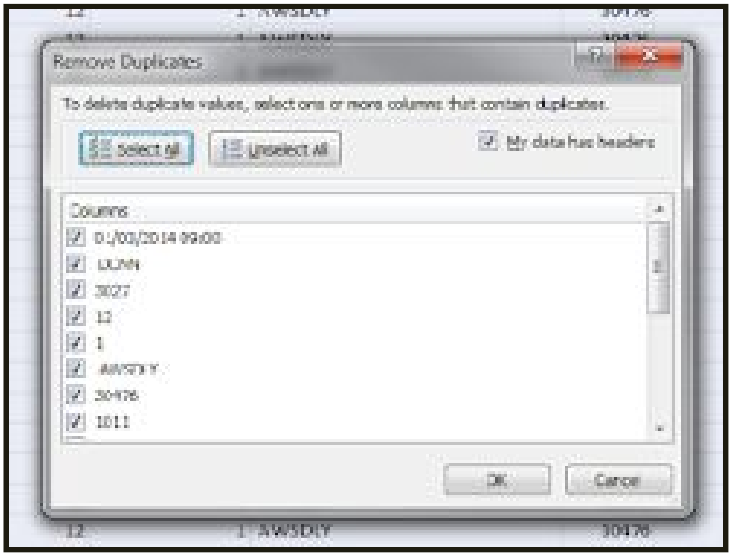
Consequently, further steps should be undertaken to try and prepare the data ahead of any analysis and to understand the limitations of the data presented. It may also be possible that true duplicate lines may have appeared in the archive.

Step 1 - remove 'duplicate' lines:

To remove the duplicates select all the data and select Data tab, in the Data Tools group, click Remove Duplicates (Highlighted in yellow).



- To quickly clear all columns, click Unselect All.



2. Click OK. A message is displayed indicating how many duplicate values were removed and how many unique values remain, or if no duplicate values were removed. Click OK.

Step 2 - Version Number and Quality control

The data arrives at the Met Office and it undergoes quality control to check that the data are correct and consistent with the surrounding data points.

As noted above, 'version 1' data should contain the latests state of the data line as held within the MIDAS system, thus removing version 0 lines will ensure that the original values, i.e. before any changes have been applied, are removed. However, this may still result in more than one 'version 1'

entry. Here the record time stamp should act as a sufficient delineator to find the latest entry.

The Met Office MIDAS system undertakes systematic quality control checks on the data automatically as well as occasional manual alterations. For each data column in the data there should be an associated QC entry - denoted by the _q and _j values. Use of these flags can help to remove unwanted datalines that would otherwise contaminate data analysis. However, the values to use may not be immediately obvious:

Quality control (QC) guide

Quality control is when the data are checked to make sure it is correct and consistent with the surrounding data points.

These data are displayed with a QC code composed of up to 5 digits e.g. 1006, 15006 or 6

Each of the 5 digits represents an entry in the 'MESQL' quality control flagging entries, filled in from the right. The letters 'MESQL' stand for the following types of QC information:

M- marker
E- estimate
S- status
Q-query
L-level

Not all types of QC entries are required and, as stated above, the value should be read from the RIGHT. For example:

Example 1

M	E	S	Q	L
	1	0	0	6

Example 2

M	E	S	Q	L
				6

The numbers represent different information for each of the QC types and can be found in detail on teh following page:

https://data.ceda.ac.uk/badc/ukmo-midas/metadata/doc/QC_J_flags.html

For example, a QC of 1006 has the following values from the QC codes:

For example

M	E	S	Q	L
	1	0	0	6

E (1) - Estimate/correction derived automatically from a program with no manual intervention

S (0)-Observed and not suspect

Q (0)-Original value is/was not queried, or no information available

L (6)-Final (or only) areal or buddy job run and queries processed

Met element name _j

This attribute is a single character code which either describes the method of measurement, or further qualifies the meteorological values. The meaning of any value depends on the element being qualified.

For example

This data below is taken from the UK Daily rainfall data

```
005349, RAIN, 1920-01-01 00:00, 1, DLY3208, 900, 1, 310, 1001, 14, , 0, , , D
```

This therefore means we need to look under the Precipitation codes.
The **D** therefore implies the rainfall has been converted from inches.

Full listing is available in the link below
https://data.ceda.ac.uk/badc/ukmo-midas/metadata/doc/QC_J_flags.html:

State Indicators

A state indicator is an attribute (rec_st_ind) of each table that is used to describe the current stage in the life of a particular record, from creation to deletion.

Full listing is available in the link below:
https://data.ceda.ac.uk/badc/ukmo-midas/metadata/doc/state_indicators.html